Computer Arithmetic for Probability Distribution Variables

Weiye Li, liw@lanl.gov
James Mac Hyman, jh@lanl.gov

Computational uncertainties are unavoidable in numerical calculations. The generation and propagation of uncertainties in the initial conditions, data and the constants in mathematical model can have serious implications in the reliability of the simulation and the decisions being made based on the simulation. These uncertainties can be quantified by probability distributions and the correlations or dependency relationships between the variables.

Monte Carlo is a classical approach in handling probabilistic uncertainty and it is still widely used. However, it becomes less powerful when encountering the uncertainties that have unknown dependency relationships or distributions that are not fully specified. Various non-Monte Carlo methods have been developed to deal with unknown dependency relationships and imprecise probabilities since 1960s. Interval arithmetic is one of the main approaches, in which intervals are considered independent in order to bound all the possible solutions. The bounds obtained by this are usually pessimistic when dependency relationships exist.

It is important to know the dependency relationships between variables in computation in order to obtain tight bounds for the possible results. Various researchers have been working on approaches concerning dependency, and progresses have been made in gaining tight bounds. However, no approach that calculates the sharp bounds had been found until our recent discovery of *Probability Distribution Variable Arithmetic*, or *PDV Arithmetic*, which extends the interval arithmetic approach with the exclusive feature of *complete dependency tracking* throughout computation.

A PDV is a random variable and is characterized by its generalized probabilistic discretization, which is a set of pairs of bins and probabilities. Using generalized probabilistic discretization is required because computers can only store discrete quantities. It is also an effective way to represent uncertainty, especially when the probability distributions are not fully specified from lack of sufficient information. In view of the fact that different random variables may have the same generalized probabilistic discretization, a PDV may be considered as a family of random variables that have the same generalized probabilistic discretization. In this point of view, every random variable in the family is a representative of the PDV.

In a computation, all variables involved are put into two category: input variables and derived variables. The latter is derived from the former via deterministic function expression. Thus, the dependency relationship between two derived variables can be well-defined by the relationship between the two pre-image sets, which are the sets of the input variables that define the derived variables. The extent of dependency can be represented by how much the two pre-image sets overlap.

Every binary operation between two PDVs is turned into the same operation between the corresponding bins. Dependency tracking requires that not every arbitrary pair of bins can be grouped to be operated on. We know that a bin of a derived PDV is fully determined by some bins of the input PDVs. To see whether two bins from the two PDVs can be paired, one needs to compare their corresponding bins from the input PDVs and justify whether they are compatible. Only the bins with compatible bins from the input PDVs can be grouped and then operated on using interval arithmetic. Under the assumption that all input PDVs are probabilistically independent, the associated probabilities of the bins can be computed.

PDV Arithmetic is formulated based on the above ideas. It can be proved that the bounds calculated by using PDV Arithmetic include all the possible solutions and these bounds converge to the sharp bounds as the widths of the refinements of the input PDVs tends to 0. Sensitivity analysis shows that these bounds are stable in the sense that small perturbations of the input bins do not affect the bounds significantly provided that no singularity occurs in the computation.

As an application, we implement PDV Arithmetic in Fortran 77 by including PDV as a basic data type. A software package PDVFOR77 that includes a preprocessor written in Perl and a subroutine library in Fortran 77 enables the user to write PDV in a program as simple as writing real and integer. Every statement in the program involving a PDV data type is parsed into a sequence of subroutine calls that implement PDV Arithmetic.

An example about the eigenvalues of a 2×2 random matrix are demonstrated in the following three figures. The first figure illustrates the dependency relationship between the two eigenvalues, and the second and third figures compare PDV Arithmetic with Monte Carlo simulation.

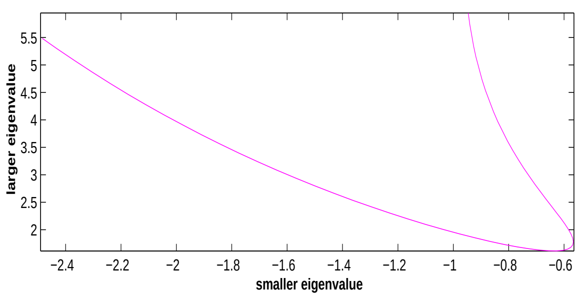
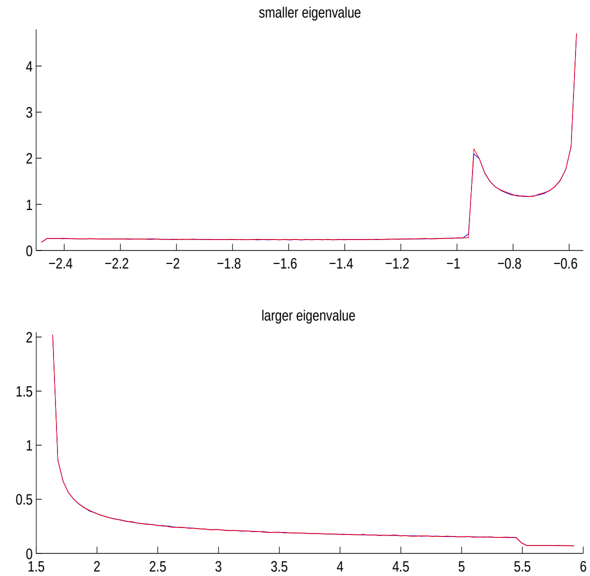
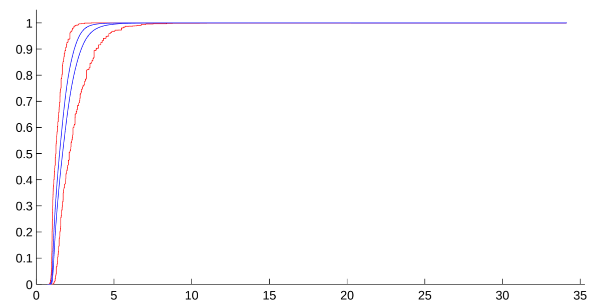


Illustration of dependency between the two eigenvalues of matrix $\left((a,c)^T,(b,d)^T\right)$ *where* $a = 3d^2+1$, $c = 2a-1$, $b = 0.5c/a+0.5$, $d \in [-1,1]$ *and d is uniformly distributed.*

Another example is given in the last figure to illustrate the probability distribution bounds (p-box, in red) calculated by PDV Arithmetic for the output of system $y = (a+b)^a$, where a belongs to 3 independent intervals $[0.8,1.0]$, $[0.5,0.7]$, $[0.1,0.4]$, and $\ln b$ follows $N(\mu,\sigma)$ where μ belongs to 3 independent intervals $[0.6,0.8]$, $[0.1,0.4]$, $[0.0,1.0]$, and σ belongs to 3 independent intervals $[0.4,0.5]$, $[0.25,0.35]$, $[0.1,0.2]$. In contrast, the trend of a special probability distribution (refined p-box, in blue) is calculated with an additional assumption that a, μ and σ are uniformly distributed in each of the above intervals.



Comparisons between PDV Arithmetic (red line) and Monte Carlo simulation (blue line) for the probability density functions of the eigenvalues of the matrix in Figure 1.



Probability distribution bounds for the output of $y = (a+b)^a$.

The details about PDV Arithmetic can be found in [1]. The PDVFOR77 package and the above paper can also be found at `http://math.lanl.gov/~liw/`.

Acknowledgments

Los Alamos Report LAUR-02-7458. Funded by the US Department of Energy, under contract W-7405-ENG-36 and the DOE Office of Science ASCR Program in Applied Mathematics.

References

[1] Li, W. and Hyman, J.M., Computer Arithmetic for Probability Distribution Variables, *Reliability Engineering and System Safety*, in print, 2004.